

*Proceedings of the 27th International Conference on Probabilistic, Combinatorial  
and Asymptotic Methods for the Analysis of Algorithms  
Kraków, Poland, 4-8 July 2016*

# Locally Restricted Sequential Structures and Runs of a Subcomposition in Integer Compositions

Edward A. Bender<sup>1</sup>, Zhicheng Gao<sup>2†</sup>

<sup>1</sup>*Department of Mathematics, University of California, San Diego, La Jolla, CA 92093-0112*

<sup>2</sup>*School of Mathematics and Statistics, Carleton University, Ottawa, Ontario K1S5B6*

---

We study part sizes of supercritical locally restricted sequential structures. This extends previous results about locally restricted integer compositions and part sizes in smooth supercritical compositional structures. Applications are given for runs of subcompositions. The problems are formulated as enumerating directed walks in sized infinite digraphs and the proofs depend heavily on earlier results by Bender and Canfield about infinite transfer matrices.

**Keywords:** sequential structure, local restriction, infinite transfer matrix, composition, run

---

## 1 Introduction

In [4] part sizes of compositional structures were studied. It was shown that if the composition is smooth supercritical then the numbers of parts of large sizes are asymptotically Poisson and an asymptotic expression was obtained for the expected value of the maximum part size which is accurate up to  $o(1)$ . In [3], part sizes of locally restricted nearly free integer compositions were studied. We will extend some of the major results of [3, 4] to locally restricted supercritical sequential structures. Runs of a single letter/part in words/compositions have been studied extensively; see e.g. [4, 6, 12]. In locally restricted structures, such as Carlitz compositions and Smirnov words, runs of a single part may not be allowed, and hence it is natural to consider runs of a substructure. Our main results on large part size distributions will be applied to maximum run length of a subcomposition in several classes of locally restricted compositions. Our approach follows that in [2, 3] using directed walks in sized infinite digraphs and properties of the corresponding infinite transfer matrices. Our proofs rely heavily on results in [2, 3] about infinite transfer matrices and large part size distributions.

## 2 Definitions

**Definition 1 (Sequential structures)** *Let  $\mathcal{P}$  be a class of combinatorial structures, called parts.*

---

<sup>†</sup>Research supported by NSERC

- Each part  $p$  has a positive integer size, denoted  $|p|$ .
- We use  $\mathcal{P}_n$  to denote the set of parts of size  $n$  and assume  $P_n := |\mathcal{P}_n| < \infty$  for each  $n \in \mathbb{N}$ .
- For each integer  $k \geq 0$ , we denote the class consisting of sequences of  $k$  parts by

$$\text{SEQ}_k(\mathcal{P}) := \{p_1 p_2 \dots p_k : p_j \in \mathcal{P}\}.$$

Thus  $\text{SEQ}_0(\mathcal{P})$  contains only the empty sequence which has size and length 0. **Throughout this paper**  $\varepsilon$  denotes the empty sequence.

- Let  $\text{SEQ}_{<k}(\mathcal{P})$  denote the class of sequences of at most  $k - 1$  parts. Let  $\text{SEQ}(\mathcal{P}) := \cup_{k \geq 0} \text{SEQ}_k(\mathcal{P})$ , the set of all sequences.
- If  $\mathbf{a} = p_1 p_2 \dots p_k \in \text{SEQ}(\mathcal{P})$ , then the length of  $\mathbf{a}$  is  $k$  and we write  $\text{len}(\mathbf{a}) = k$ . The size of  $\mathbf{a}$  is

$$|\mathbf{a}| = |p_1| + |p_2| + \dots + |p_k|$$

and the distance from  $p_i$  to  $p_j$  is  $|i - j|$ .

**Example 1 (Generalized compositions)** When studying  $\text{SEQ}(\mathcal{P})$ , only the values of  $P_n$  are important. Thus we could think of  $\mathcal{P}$  as parts in a generalized composition where the part  $n$  comes in  $P_n$  “colors”, or whatever you choose to call them. If  $P_n = 0$ , there are no parts of size  $n$ . Thus, ordinary compositions correspond to  $P_n = 1$  for all  $n$  and words on a  $k$ -letter alphabet correspond to  $P_1 = k$  and  $P_n = 0$  for  $n > 1$ . When parts of size  $n$  occur in  $n$  colors,  $P_n = n$  and we have  $n$ -colored compositions, which were studied in [1, 9, 11]. We may also consider colored compositions where a part of size  $n$  corresponds to a multiset of  $n$  colored balls with  $N$  colors available. Here we have  $P_n = \binom{n+N-1}{N-1}$ .  $\square$

Let  $\mathcal{A} \subseteq \text{SEQ}(\mathcal{P})$ . It is called locally restricted if the parts of a structure in  $\mathcal{A}$  within a fixed distance satisfy certain restrictions. Locally restricted integer compositions were studied in [2], where local restrictions are defined in terms of local restriction functions. The function was then used to construct a digraph. In this paper, we define local restrictions directly in terms of the digraph. Readers wishing to see the connection between local restriction functions and the digraph should consult [2].

**Definition 2 (Locally restricted structures)** Let  $m \in \mathbb{N}$ ,  $\mathcal{S}, \mathcal{F} \subseteq \text{SEQ}_{<m}(\mathcal{P})$  and  $\mathcal{R} \subseteq \text{SEQ}_m(\mathcal{P})$ . The integer  $m$  is called the span of the locally restricted class of structures associated with the digraph  $D$  which has vertex set  $V(D) = \mathcal{S} \cup \mathcal{R} \cup \mathcal{F}$ . If  $p \in \mathcal{S} \cap \mathcal{F}$ , we allow two copies as separate vertices in the digraph. For convenience, we introduce two copies of the empty sequence, denoted by  $\varepsilon_s$  and  $\varepsilon_f$ , such that  $\varepsilon_s \in \mathcal{S}$  and  $\varepsilon_f \in \mathcal{F}$ . Suppose  $D$  satisfies the following conditions.

- (a) There is an arc from  $\varepsilon_s$  to every other vertex in  $\mathcal{S}$ , and at least one arc from  $\mathcal{S}$  to  $\mathcal{R}$ .
- (b) There is an arc to  $\varepsilon_f$  from every other vertex in  $\mathcal{F}$ , and at least one arc from  $\mathcal{R}$  to  $\mathcal{F}$ .
- (c) The sub-digraph  $D_{\mathcal{R}}$  of  $D$  induced by  $\mathcal{R}$  is strongly connected,  $|\mathcal{R}| \geq 2$ , and the sub-digraph of  $D$  induced by  $\mathcal{S} \cup \mathcal{F}$  contains no directed cycle.

The vertices of  $\mathcal{R}$  will be called the recurrent vertices, and the vertices of  $\mathcal{S}$  and  $\mathcal{F}$  will be called the start vertices and finish vertices, respectively. If a part  $p \in \mathcal{P}$  appears in some recurrent vertex, then  $p$  is called a recurrent part.

Let  $\mathcal{W}$  denote the set of all directed walks in  $D$  from  $\varepsilon_s$  to  $\varepsilon_f$ . We use  $\text{SEQ}(\mathcal{P}; D)$  to denote the class of all structures of the concatenation form  $\mathbf{v}_1 \mathbf{v}_2 \cdots \mathbf{v}_j$ , where  $\varepsilon_s \mathbf{v}_1 \mathbf{v}_2 \cdots \mathbf{v}_j \varepsilon_f \in \mathcal{W}$ . If each element in  $\text{SEQ}(\mathcal{P}; D)$  arises from **only one** walk in  $\mathcal{W}$ , we say that  $\text{SEQ}(\mathcal{P}; D)$  is a locally restricted class associated with  $D$ .

The “only one” condition is required so that an ogf built from  $D$  will count each element in  $\text{SEQ}(\mathcal{P}; D)$  just once. The span is actually associated with the digraph. One can easily construct a digraph  $D'$  with span  $km$  for any  $k \in \mathbb{N}$  such that  $\text{SEQ}(\mathcal{P}; D') = \text{SEQ}(\mathcal{P}; D)$ .

**Definition 3 (Regular class)** A class  $\text{SEQ}(\mathcal{P}; D)$  of locally restricted structures will be called regular if it satisfies the following conditions.

- The gcd of the lengths of all directed cycles in  $D_{\mathcal{R}}$  is equal to 1.
- There is a positive integer  $k$  and vertices  $\mathbf{v}_0, \mathbf{v}_k \in \mathcal{R}$  such that  $\gcd\{m - n : m, n \in S\} = 1$ , where  $S = \{n : n = |\mathbf{v}_0| + \cdots + |\mathbf{v}_k| \text{ for some directed walk } \mathbf{v}_0 \cdots \mathbf{v}_k \text{ of length } k \text{ in } D_{\mathcal{R}}\}$ .

**Example 2 (Pattern avoidance)** Let  $B$  be a finite set of structures in  $\text{SEQ}(\mathcal{P})$ , and  $\mathcal{A}$  be the class of structures in  $\text{SEQ}(\mathcal{P})$  which don't contain any structure in  $B$ . We say that the structures in  $\mathcal{A}$  avoid the structures in  $B$  (or simply, avoid  $B$ ). We may construct the class  $\mathcal{A}$  using the following digraph. Let  $m+1$  be the maximum length of the structures in  $B$ . Let  $\mathcal{S} = \{\varepsilon_s\}$ ,  $\mathcal{F} \subset \text{SEQ}_{<m}(\mathcal{P})$  and  $\mathcal{R} \subset \text{SEQ}_m(\mathcal{P})$  be consisting of structures which avoid  $B$ . There is an arc from vertex  $\mathbf{a}$  to a vertex  $\mathbf{b}$  if and only if  $\mathbf{ab}$  avoids  $B$ . There is also an arc from  $\varepsilon_s$  to every vertex in  $\mathcal{R} \cup \mathcal{F}$ . Then  $\mathcal{A} = \text{SEQ}(\mathcal{P}; D)$ . We note that the second gcd condition in Definition 3 is not satisfied here because each recurrent vertex has size  $m$ .  $\square$

**Remark:** Words over a finite alphabet which avoid certain patterns have been studied extensively (see, e.g., [8]). Here the size of a word is the length of the word. Since  $\mathcal{P}$  is finite here, one may use the simpler transfer matrix  $T(z)$  such that  $T_{i,j}(z) = z^{|\mathbf{r}_j|}$  for recurrent vertex  $\mathbf{r}_j$  (See Definition 6 on page 8 for the transfer matrix). Consequently the ogf  $F_R(z) = \mathbf{s}^t(I - T(z))^{-1}\mathbf{f}$  is a rational function of  $z$ . We note that in this case, each recurrent vertex has size  $m$  and so  $T(z)$  is a function of  $z^m$ , and hence we may apply [2, Theorem 1] to  $T$  with  $x = z^m$ .

**Definition 4 (Generating functions and supercritical structures)** Let  $\mathcal{A} \subseteq \text{SEQ}(\mathcal{P})$ .

- Define the ordinary generating function (ogf)  $P(z) := \sum_{p \in \mathcal{P}} z^{|p|} = \sum_{n \geq 1} P_n z^n$ .
- Throughout this paper  $\rho$  is the radius of convergence of  $P(z)$ .
- If the radius of convergence of the ogf  $A(z) := \sum_{\mathbf{a} \in \mathcal{A}} z^{|\mathbf{a}|}$  is less than  $\rho$ , we call  $\mathcal{A}$  supercritical.

**Example 3 (Alternating compositions)** Alternating, or up-down, compositions are compositions in which a part is alternately greater and less than the preceding part. We set  $m = 2$ ,  $\mathcal{P} = \mathbb{N}$ ,  $\mathcal{S} = \text{SEQ}_{<2}(\mathcal{P})$ ,  $\mathcal{F} = \text{SEQ}_{<2}(\mathcal{P}) \setminus \{1\}$ , and  $\mathcal{R} = \{i, j \mid i > j\}$ . The empty sequence in  $\mathcal{S}$  (resp.  $\mathcal{F}$ ) connects to every element in  $\mathcal{R}$ . There is an arc from  $i \in \mathcal{S}$  to  $j \in \mathcal{F}$  whenever  $i < j$  or  $i$  is the empty sequence. The arcs having at least one end in  $\mathcal{R}$  should be fairly easy to see. Every alternating composition, including the empty one, is associated with a unique directed walk from  $\varepsilon_s$  to  $\varepsilon_f$ . It is known [2] that the class of alternating compositions is regular and supercritical.  $\square$

**Definition 5 (Asymptotically free set)** Let  $\mathcal{A} = \text{SEQ}(\mathcal{P}; D)$  be a class of locally restricted structures with  $|\mathcal{P}| = \infty$  and let  $\mathcal{L}$  be an infinite set of recurrent parts. If  $\mathbf{a} = p_1 \dots p_k$  and  $t = \min(k, m - 1)$ , define  $s(\mathbf{a}) := p_1 \dots p_t$  and  $f(\mathbf{a}) := p_{k-t+1} \dots p_k$ . Suppose there is a function  $g : \text{SEQ}_{<m}(\mathcal{P}) \times \text{SEQ}_{<m}(\mathcal{P}) \rightarrow \mathbb{N}$  such that the following holds. If  $\mathbf{a}l\mathbf{z} \in \text{SEQ}(\mathcal{P}; D)$  with  $l \in \mathcal{L}$  and  $|l| \geq g(f(\mathbf{a}), s(\mathbf{z}))$ , then  $\mathbf{a}l'\mathbf{z} \in \text{SEQ}(\mathcal{P}; D)$  whenever  $l' \in \mathcal{L}$  and  $|l'| \geq g(f(\mathbf{a}), s(\mathbf{z}))$ . Then we call  $\mathcal{L}$  asymptotically free.

What this says is that if  $l$  is large enough as determined by parts closer than distance  $m$ , then it can be replaced by any large enough part in  $\mathcal{L}$ .

**Example 4 ( $k$ -Carlitz compositions)** A  $k$ -Carlitz composition is a composition in which each part is different from each of the preceding  $m$  parts, i.e.  $p_i \neq p_j$  whenever  $|i - j| \leq m$ . (Carlitz compositions, for which adjacent parts differ, is the case  $k = 1$ . They have been studied extensively. See for example [10].) They are regular supercritical structures and a digraph  $D$  can be constructed with any  $m \geq k$ . Let  $\mathcal{P} = \mathbb{N}$  and  $\mathcal{S} = \{\varepsilon_s\}$ . Let  $\mathcal{R}$  (respectively  $\mathcal{F}$ ) be all compositions in  $\text{SEQ}_m(\mathcal{P})$  (respectively  $\text{SEQ}_{<m}(\mathcal{P})$ ) which are  $k$ -Carlitz. There is an arc from  $\varepsilon_s$  to all vertices in  $\mathcal{R} \cup \mathcal{F}$ . There is an arc from  $\mathbf{a} \in \mathcal{R}$  to  $\mathbf{b} \in \mathcal{R} \cup \mathcal{F}$  whenever  $\mathbf{a}\mathbf{b}$  is an  $k$ -Carlitz composition. Since a part that is larger than all parts within distance  $k$  can be replaced by any larger part, the set  $\mathbb{N}$  is asymptotically free in  $k$ -Carlitz compositions.  $\square$

### 3 Main Results

Let  $r$  be the radius of convergence of the ogf for  $\text{SEQ}(\mathcal{P}; D)$ . In the following, all logarithms will be to the base  $1/r$ . Our main results are the following.

**Theorem 1** Let  $\mathcal{A} = \text{SEQ}(\mathcal{P}; D)$  be a regular supercritical class of locally restricted structures, and  $\mathcal{L} = \{l_1, l_2, \dots\}$  be an asymptotically free set. Assume  $|l_1| < |l_2| < \dots$ . Select a structure  $\mathbf{a}$  uniformly at random from  $\mathcal{A}_n$ . Let  $\zeta_k(n)$  be the number of occurrences of  $l_k$  in  $\mathbf{a}$ . The following are true.

- (a)  $|\mathcal{A}_n| = Ar^{-n} (1 + O(e^{-\delta n}))$  for some positive constants  $r$ ,  $A$  and  $\delta$ .
- (b) The distribution of  $\zeta_k(n)$  is asymptotically normal with mean and variance asymptotically proportional to  $n$ .
- (c) The limit

$$v_k = \lim_{n \rightarrow \infty} \frac{\mathbb{E}(\zeta_k(n))}{n} \quad (1)$$

exists, and  $v_k \sim Cr^{|l_k|}$  as  $k \rightarrow \infty$ , for some positive constant  $C$ .

- (d) Suppose there is a function  $\omega_1(n) \rightarrow \infty$  such that  $\{|l_k| : k \geq 1\} \cap [\log n - \omega_1(n), n]$  is not empty for all sufficiently large  $n$ . Then there is a function  $\omega_2(n) \rightarrow \infty$  such that the random variables  $\{\zeta_k(n) : \log n - \omega_2(n) \leq |l_k| \leq n\}$  are asymptotically independent Poisson random variables with means  $\mu_k = Cnr^{|l_k|}$ .

The distribution of large part sizes in general supercritical compositional structures has been studied extensively. Some latest results can be found in [4]. We will convert runs of a given subcomposition into free parts in a related class of locally restricted structures and apply Theorem 1(d) to derive the following.

**Theorem 2** Let  $\mathbf{c}$  be a given composition such that  $\mathbf{c}$  cannot be written in the form  $\mathbf{c} = \mathbf{xyx}$ . Let  $\mathcal{A} = \text{SEQ}(\mathbb{N}; D)$  be a regular supercritical class of locally restricted compositions with span  $m = \text{len}(\mathbf{c})$ . Assume that  $\mathbf{c}$  is a recurrent vertex and the digraph  $D$  contains the arc from  $\mathbf{c}$  to itself. Let  $C, r, \mathbf{a}$  and  $\zeta_k$  be as defined in Theorem 1,  $R_n$  be the maximum run length of  $\mathbf{c}$  in  $\mathbf{a}$ , and  $g_n(k)$  be the probability that  $\mathbf{a}$  contains exactly  $k$  runs of  $\mathbf{c}$  of length  $R_n$ . Let  $\gamma \doteq 0.577216$  be Euler's constant and define

$$P_k(x) = \frac{\log e}{|\mathbf{c}|} \sum_{\ell \neq 0} \Gamma\left(k + \frac{2i\pi\ell \log e}{|\mathbf{c}|}\right) \exp\left(\frac{-2i\ell\pi \log x}{|\mathbf{c}|}\right). \quad (2)$$

Then

- (a)  $\mathbb{P}(R_n < k) \sim \exp\left(-\frac{Cn}{1-r^{|\mathbf{c}|}} r^{k|\mathbf{c}|}\right)$ .
- (b)  $\mathbb{E}(R_n) = \frac{1}{|\mathbf{c}|} \log \frac{Cn}{1-r^{|\mathbf{c}|}} + \frac{\gamma \log e}{|\mathbf{c}|} - \frac{1}{2} - P_0\left(\frac{Cn}{1-r^{|\mathbf{c}|}}\right) + o(1)$ .
- (c)  $g_n(k) = \frac{(1-r^{|\mathbf{c}|})^k}{k!} P_k\left(\frac{Cn}{1-r^{|\mathbf{c}|}}\right) + \frac{(1-r^{|\mathbf{c}|})^k \log e}{k|\mathbf{c}|} + o(1)$ .

**Corollary 1** Theorem 2 holds for the following classes of compositions.

- (a) All compositions for any given composition  $\mathbf{c}$ , where  $r = 1/2$  and  $C = \frac{1}{2}(1-2^{-|\mathbf{c}|})^2$ . In particular,  $C = 1/8$  when  $\mathbf{c} = 1$ , which gives Gafni's result [6].
- (b) Carlitz compositions for any given Carlitz composition  $\mathbf{c}$ , where  $r \doteq 0.571350$  is the smallest positive number satisfying  $\sum_{j \geq 1} \frac{r^j}{1+r^j} = 1$ . In particular, when  $\mathbf{c} = ab$  with  $a \neq b$ , we have

$$C = \frac{(1-r^{a+b})^2}{(1+r^a)(1+r^b)} \frac{1}{\sum_{j \geq 1} \frac{jr^j}{(1+r^j)^2}}.$$

- (c)  $k$ -Carlitz compositions for any given  $k$ -Carlitz composition  $\mathbf{c}$ .
- (d) Alternating compositions for any given alternating composition  $\mathbf{c}$ , where  $r \doteq 0.6363$ .
- (e)  $n$ -color compositions (defined in Example 1) for any given  $n$ -color composition  $\mathbf{c}$ , where  $r = \frac{3-\sqrt{5}}{2}$ .
- (f) Colored compositions with  $P_n = \binom{n+N-1}{N-1}$  (defined in Example 1) for any given colored composition  $\mathbf{c}$ , where  $r = 1 - 2^{-1/N}$ .

## 4 Converting Runs into Run Parts

The basic idea in the proof of Theorem 2 is to replace a  $k$ -long run of a given subcomposition  $\mathbf{c}$  with a new part  $\bar{k}$  with  $|\bar{k}| = k|\mathbf{c}|$ , and then apply Theorem 1(d) to a new class  $\mathcal{A}'$  of locally restricted structures with parts in  $\mathbb{N}$  as well as *run parts*  $\bar{k}$  for  $k \geq 1$ . Let  $\theta$  denote this replacement operation. For example, if  $\mathbf{c} = 12$  and  $\mathbf{a} = 12123122121$ , then  $\theta(\mathbf{a}) = 23\bar{1}2\bar{1}1$ .

**Example 5 (Runs in unrestricted compositions)** Let  $\mathcal{A}$  be the class of all compositions, and  $\mathbf{c}$  be a given composition with  $\text{len}(\mathbf{c}) = m$  such that it cannot be written in the form  $\mathbf{c} = \mathbf{xyx}$ . The digraph  $D'$  for  $\mathcal{A}' = \theta(\mathcal{A})$  is defined as follows.

- $\mathcal{S}(D') = \{\varepsilon_s\}$ .
- $\mathbf{a}' \in \mathcal{F}(D')$  if and only if  $\mathbf{a}' \in \text{SEQ}_{<m}(\mathcal{P})$ ,  $\mathbf{a}'$  does not contain two consecutive run parts.
- $\mathbf{r}' \in \mathcal{R}(D')$  if and only if  $\mathbf{r}' \in \text{SEQ}_m(\mathcal{P})$ ,  $\mathbf{r}' \neq \mathbf{c}$ ,  $\mathbf{r}'$  does not contain two consecutive run parts.
- There is an arc from a vertex  $\mathbf{a}'$  to a vertex  $\mathbf{b}'$  if and only if  $\mathbf{a}'\mathbf{b}'$  does not contain  $\mathbf{c}$  and does not contain two consecutive run parts.

□

Let  $\mathbf{c}$  be a given composition which cannot be written in the form  $\mathbf{c} = \mathbf{xyx}$ . It is easy to see that copies of  $\mathbf{c}$  in a compositions  $\mathbf{a}$  cannot overlap with each other. Define  $\mathcal{P} = \mathbb{N} \cup \{\bar{k} : k \geq 1\}$  with  $|\bar{k}| = k|\mathbf{c}|$ . We have the following

**Proposition 1** Let  $\mathcal{A} = \text{SEQ}(\mathcal{P}; D)$  be a given regular supercritical class of locally restricted compositions with span  $m$  such that  $\mathcal{S}(D) = \{\varepsilon_s\}$ . Let  $\mathbf{c} \in \mathcal{R}(D)$ . Assume that  $D$  contains an arc from  $\mathbf{c}$  to itself and  $\mathbf{c}$  cannot be written in the form  $\mathbf{c} = \mathbf{xyx}$ . Let  $\mathcal{A}' = \theta(\mathcal{A})$ . Then

- $\mathcal{A}' = \text{SEQ}(\mathcal{P}; D')$  is a regular supercritical class with  $\mathcal{P} = \mathbb{N} \cup \{\bar{k} : k \geq 1\}$  and some digraph  $D'$ .
- For each  $\mathbf{a} \in \mathcal{A}$ ,  $|\theta(\mathbf{a})| = |\mathbf{a}|$ , and  $\theta$  is a bijection between  $\mathcal{A}_n$  and  $\mathcal{A}'_n$ .
- For each  $\mathbf{a} \in \mathcal{A}$ , Let  $R(\mathbf{a})$  be the maximum run length of  $\mathbf{c}$  in  $\mathbf{a}$ , and  $M(\mathbf{a})$  be the maximum value of  $k$  such that  $\bar{k}$  appears in  $\theta(\mathbf{a})$ . We have  $R(\mathbf{a}) = M(\mathbf{a})$ .

*Proof:* (a) We define the digraph  $D'$  as follows.

- $\mathcal{S}(D') = \{\varepsilon_s\}$ .
- $\mathbf{a}' \in \mathcal{F}(D')$  if and only if  $\mathbf{a}' \in \text{SEQ}_{<m}(\mathcal{P})$ ,  $\mathbf{a}'$  does not contain two consecutive run parts and  $\theta^{-1}(\mathbf{a}')$  appears at the end of a composition in  $\mathcal{A}$ .
- $\mathbf{r}' \in \mathcal{R}(D')$  if and only if  $\mathbf{r}' \in \text{SEQ}_m(\mathcal{P})$ ,  $\mathbf{r}' \neq \mathbf{c}$ ,  $\mathbf{r}'$  does not contain two consecutive run parts, and  $\theta^{-1}(\mathbf{r}')$  appears in a composition in  $\mathcal{A}$ .
- There is an arc from a vertex  $\mathbf{a}'$  to a vertex  $\mathbf{b}'$  if and only if  $\mathbf{a}'\mathbf{b}'$  does not contain two consecutive run parts, does not contain  $\mathbf{c}$ , and  $\theta^{-1}(\mathbf{a}'\mathbf{b}')$  appears in a composition in  $\mathcal{A}$ .

It can be verified that  $\mathcal{A}' = \text{SEQ}(\mathcal{P}; D')$  is regular supercritical.

(b) This follows immediately from the definition of  $\theta$ .

(c) Since the copies of  $\mathbf{c}$  in  $\mathbf{a}$  cannot overlap with each other,  $\theta$  induces a bijection between the set of run parts in  $\theta(\mathbf{a})$  and the set of runs in  $\mathbf{a}$ , which implies  $R(\mathbf{a}) = M(\mathbf{a})$ .  $\square$

**Example 6 (Runs of a subcomposition in  $m$ -Carlitz compositions)** Let  $\mathbf{c}$  be a sequence of  $m$  distinct integers. The digraph  $D'$  for  $\mathcal{A}' = \theta(\mathcal{A})$  is defined as follows.

- $\mathcal{S}(D') = \{\varepsilon_s\}$ .
- $\mathbf{a}' \in \mathcal{F}(D')$  if and only if  $\mathbf{a}' \in \text{SEQ}_{<m}(\mathcal{P})$ ,  $\mathbf{a}'$  does not contain two consecutive run parts and  $\theta^{-1}(\mathbf{a}')$  is  $m$ -Carlitz.
- $\mathbf{r}' \in \mathcal{R}(D')$  if and only if  $\mathbf{r}' \in \text{SEQ}_m(\mathcal{P})$ ,  $\mathbf{r}' \neq \mathbf{c}$ ,  $\mathbf{r}'$  does not contain two consecutive run parts, and  $\theta^{-1}(\mathbf{a}')$  is  $m$ -Carlitz.
- There is an arc from a vertex  $\mathbf{a}'$  to a vertex  $\mathbf{b}'$  if and only if  $\mathbf{a}'\mathbf{b}'$  does not contain  $\mathbf{c}$ , does not contain two consecutive run parts, and  $\theta^{-1}(\mathbf{a}'\mathbf{b}')$  is  $m$ -Carlitz.

$\square$

**Example 7 (Runs of a subcomposition in alternating compositions)** Let  $m \geq 2$ ,  $\mathbf{c} = c_1c_2 \cdots c_m$  be an alternating composition such that  $\mathbf{c}^2$  is also alternating. The digraph  $D'$  for  $\mathcal{A}' = \theta(\mathcal{A})$  is defined as follows.

- $\mathcal{S}(D') = \{\varepsilon_s\}$ .
- $\mathbf{a}' \in \mathcal{F}(D')$  if and only if  $\mathbf{a}' \in \text{SEQ}_{<m}(\mathcal{P})$ ,  $\mathbf{a}'$  does not contain two consecutive run parts and  $\theta^{-1}(\mathbf{a}')$  is alternating.
- $\mathbf{r}' \in \mathcal{R}(D')$  if and only if  $\mathbf{r}' \in \text{SEQ}_m(\mathcal{P})$ ,  $\mathbf{r}' \neq \mathbf{c}$ ,  $\mathbf{r}'$  does not contain two consecutive run parts, and  $\theta^{-1}(\mathbf{a}')$  is alternating.
- There is an arc from a vertex  $\mathbf{a}'$  to a vertex  $\mathbf{b}'$  if and only if  $\mathbf{a}'\mathbf{b}'$  does not contain  $\mathbf{c}$ , does not contain two consecutive run parts, and  $\theta^{-1}(\mathbf{a}'\mathbf{b}')$  is alternating.

$\square$

## 5 Outline of Proofs

The proofs are essentially the same as those in [2, 3]. In particular, we will make use of the infinite transfer matrix.

**Definition 6 (Transfer matrix)** Let  $\text{SEQ}(\mathcal{P}; D)$  be a class of locally restricted structures as in Definition 2. Let  $\mathbf{r}_1, \mathbf{r}_2, \dots$  be an ordered list of vertices in  $\mathcal{R}$ . We define the transfer matrix  $T(z)$  such that the  $(i, j)$ th entry of  $T(z)$  is  $T_{i,j}(z) = z^{|\mathbf{r}_i|+|\mathbf{r}_j|}$  if there is an arc in  $D_{\mathcal{R}}$  from  $\mathbf{r}_i$  to  $\mathbf{r}_j$ ; otherwise  $T_{i,j}(z) = 0$ .

The *weight* of an arc  $(\mathbf{v}, \mathbf{w})$  in  $D$  is  $z^{|\mathbf{v}|+|\mathbf{w}|}$ . We define the weight of a directed walk in  $D$  to be the product of the weights of all the arcs in the walk. Let  $F_R(z)$  be the ogf for structures in  $\text{SEQ}(\mathcal{P}; D)$  containing at least one recurrent vertex. It is not difficult to see that  $F_R(z^2)$  is the sum of weights of all directed walks in  $\mathcal{W}$  containing at least one recurrent vertex. We may express  $F_R(z)$  in terms of  $T(z)$ , the *start vector*  $\mathbf{s}(z)$  and the *finish vector*  $\mathbf{f}(z)$ , which are defined as follows. The  $i$ th component of  $\mathbf{s}(z)$  is the sum of weights of all directed walks from  $\varepsilon_s$  to  $\mathbf{r}_i$ , and the  $j$ th component of  $\mathbf{f}(z)$  is the sum of weights of all directed walks from  $\mathbf{r}_j$  to  $\varepsilon_f$ . Since  $T_{i,j}^k(z)$  is the sum of weights of all directed walks of length  $k$  from  $\mathbf{r}_i$  to  $\mathbf{r}_j$ , we have

$$F_R(z^2) = \mathbf{s}(z)^t \sum_{k \geq 0} T^k(z) \mathbf{f}(z).$$

The following lemma summarizes the results which are used in the proof of our Theorems 1 and 2. These results are simple extensions of the corresponding results from [2] for locally restricted compositions.

**Lemma 1** *Let  $\mathcal{A} = \text{SEQ}(\mathcal{P}; D)$  be a regular class of locally restricted structures.*

- (a) *Let  $r$  be the radius of convergence of the generating function  $F_R(z)$ . Then  $r < 1$  and it is a simple pole of  $F_R(z)$ . Moreover  $F_R(z)$  has no other singularity in  $|z| \leq 1$ .*
- (b) *Let  $p \in \mathcal{P}$  be a recurrent part and let  $X_n$  be the number of occurrences of  $p$  in a random structure in  $\mathcal{A}_n$ . There are constants  $C_i > 0$  such that*

$$\Pr(X_n < C_1 n) < C_2(1 + C_3)^{-n} \text{ for all } n.$$

- (c) *Let  $\mathbf{c}$  be a given structure in  $\mathcal{A}$ . There is a constant  $B$  such that the probability that  $\mathbf{c}$  occurs in a random structure in  $\mathcal{A}_n$  is at most  $Bnr^{|\mathbf{c}|}$ .*

*Proof of Lemma 1:* Part (a) is lifted from [2, Theorem 2] and its proof remains exactly the same.

Parts (b) and (c) are lifted from [2, Lemma 1] and their proofs remain exactly the same.  $\square$

*Proof of Theorem 1:* Theorem 1(a) follows from Lemma 1 as shown in [2]. Now the proof of Theorem 1(b–d) is essentially the same as that of [2, Theorem 1], using Lemma 1 above, and Lemmas 3, 4 and 5 from [2].  $\square$

**Lemma 2** *Let  $\mathcal{A} = \text{SEQ}(\mathbb{N}; D)$  be a class of locally restricted compositions (or colored compositions defined in Example 1). Then  $\mathcal{A}$  is supercritical.*

*Proof:* Let  $G(z)$  be the ogf for all compositions in  $\mathcal{A}$  which don't contain any recurrent vertex. Since each composition counted by  $G(z)$  has at most  $K$  parts for some fixed integer  $K$ , we have  $G(z) \leq \sum_{j=0}^K P(z)^j$  (coefficient-wise). Since the radius of convergence of  $P(z)$  is 1, the radius of convergence of  $G(z)$  is at least 1.  $\square$

*Proof of Theorem 2:* Let  $\mathcal{P} = \mathbb{N} \cup \{\bar{k} : k \geq 1\}$ , and  $\mathcal{A}' = \text{SEQ}(\mathcal{P}; D')$  be the new class defined in Proposition 1. Since  $\theta$  is a bijection between  $\mathcal{A}_n$  and  $\mathcal{A}'_n$ , the radius of convergence for  $\mathcal{A}'$  is the same as



that for  $\mathcal{A}$ . It is easy to see that  $\mathcal{L} = \{\bar{k} : k \geq 1\}$  is asymptotically free in  $\mathcal{A}'$ . Applying Theorem 1(d) with  $|l_j| = j|\mathbf{c}|$ , we obtain

$$\begin{aligned} \mathbb{P}(R_n < k) &= \prod_{k \leq j \leq n/|\mathbf{c}|} \mathbb{P}(\zeta_j(n) = 0) \exp(-Cnr^{j|\mathbf{c}|}) \\ &\sim \exp\left(-Cn \frac{r^{k|\mathbf{c}|}}{1 - r^{|\mathbf{c}|}}\right). \end{aligned}$$

This establishes part (a).

Using  $\mathbb{E}(R_n) = \sum_{k \geq 1} (1 - \mathbb{P}(R_n \leq k - 1))$  and the same argument as in [3, page 25] (replacing  $r$  with  $r^{|\mathbf{c}|}$ ), we obtain part (b). We remark that the  $+$  sign before  $P_0$  should be  $-$  in the expression of  $f(x)$  in [3, page 25]. The same applies to [3, Theorem 1(b,c)]

Part (c) follows from the same argument as in [3, page 27], with  $r$  being replaced by  $r^{|\mathbf{c}|}$ .  $\square$

*Proof of Corollary 1:* This follows immediately from Theorem 2. The values of  $r$  and  $C$  are computed as follows.

For part (a), it is clear  $r = 1/2$ . To obtain the value of  $C$ , we let  $F(z)$  be the ogf of compositions containing a marked  $k$ -long run of a subcomposition  $\mathbf{c}$ . We note that  $(1 - z)/(1 - 2z)$  is the ogf of all compositions, and so  $z^{|\mathbf{c}|}(1 - z)/(1 - 2z)$  is the ogf of compositions ending with (or starting with)  $\mathbf{c}$ . Hence we have

$$F(z) = \left( \frac{1 - z}{1 - 2z} - \frac{z^{|\mathbf{c}|}(1 - z)}{1 - 2z} \right)^2 z^{k|\mathbf{c}|} = (1 - z)^2 (1 - z^{|\mathbf{c}|})^2 z^{k|\mathbf{c}|} (1 - 2z)^{-2}.$$

It follows from the “transfer theorem” [5] that

$$[z^n]F(z) \sim \frac{n}{4} \left(1 - 2^{-|\mathbf{c}|}\right)^2 2^{n-k|\mathbf{c}|}.$$

Hence  $C = \frac{1}{2} (1 - 2^{-|\mathbf{c}|})^2$ . In particular,  $C = 1/8$  for the runs of 1, which gives Gafni’s result.

For part (b), the value of  $r$  can be found, for example, in [3]. The expression of  $C$  (and  $r$ ) can be derived as follows. Let  $K(z)$  be the ogf of Carlitz compositions, and  $F(z)$  be the ogf of Carlitz compositions with a marked  $k$ -long run of  $\mathbf{c} = ab$ , where  $a \neq b$ . For a given Carlitz composition  $\mathbf{v}$ , let  $K_{\mathbf{v}}(z)$  be the ogf of Carlitz compositions that start with  $\mathbf{v}$ . Then we have

$$\begin{aligned} K_a(z) &= (K(z) - K_a(z))z^a, \\ K_{ab}(z) &= z^{a+b}(K(z) - K_b(z)), \\ K_{ba}(z) &= z^{b+a}(K(z) - K_a(z)), \\ F(z) &= z^{k(a+b)}(K(z) - K_b(z) - K_{ab}(z))(K(z) - K_a(z) - K_{ba}(z)). \end{aligned} \tag{3}$$

It follows that

$$F(z) = \frac{1}{(1 + z^a)(1 + z^b)} z^{k(a+b)} (1 - z^{a+b})^2 K(z)^2. \tag{4}$$

The expression of  $K(z)$  can be obtained from (3), which gives

$$K_a(z) = \frac{z^a}{1+z^a} K(z).$$

Summing over  $a$ , and noting  $K(z) = 1 + \sum_{a \geq 1} K_a(z)$ , we obtain

$$K(z) = \frac{1}{1 - \sum_{j \geq 1} \frac{z^j}{1+z^j}}.$$

Let  $r$  be the smallest positive number satisfying  $\sum_{a \geq 1} \frac{r^a}{1+r^a} = 1$ , it is easy to see that  $r$  is a simple pole of  $K(z)$  and

$$K(z) = \frac{1}{\sum_{j \geq 1} \frac{jr^j}{(1+r^j)^2}} \frac{1}{1-z/r} + h(z),$$

where  $h(z)$  is analytic in  $|z| \leq r$ . Consequently

$$[z^n]K(z) \sim \frac{1}{\sum_{j \geq 1} \frac{jr^j}{(1+r^j)^2}} r^{-n}.$$

It follows from (4) and the “transfer theorem” that

$$\begin{aligned} [z^n]F(z) &\sim \frac{n}{(1+r^a)(1+r^b)} r^{k(a+b)} (1-r^{a+b})^2 \left( \frac{1}{\sum_{j \geq 1} \frac{jr^j}{(1+r^j)^2}} \right)^2 r^{-n}, \\ \frac{[z^n]F(z)}{[z^n]K(z)} &\sim \frac{(1-r^{a+b})^2}{(1+r^a)(1+r^b)} \frac{1}{\sum_{j \geq 1} \frac{jr^j}{(1+r^j)^2}} n r^{k(a+b)}. \end{aligned}$$

Hence

$$C = \frac{(1-r^{a+b})^2}{(1+r^a)(1+r^b)} \frac{1}{\sum_{j \geq 1} \frac{jr^j}{(1+r^j)^2}}.$$

The value of  $r$  for part (d) can be found in [3]. To obtain the value of  $r$  for part (e), we note that the corresponding part generating function is  $P(z) = z(1-z)^{-2}$ . Solving the equation  $r(1-r)^{-2} = 1$ , we obtain  $r = \frac{3-\sqrt{5}}{2}$ . To obtain the value of  $r$  for part (f), we note that the corresponding part generating function is  $P(z) = (1-z)^{-N} - 1$ . Solving the equation  $(1-r)^{-N} = 2$ , we obtain  $r = 1 - 2^{-1/N}$ .  $\square$

## 6 Discussions

In this paper, we showed how infinite transfer matrix method developed in [2] can be applied to enumerate locally restricted regular supercritical sequential structures. Poisson distribution results were derived for sizes of free parts and applications are given for runs of subcompositions in several classes of compositions. In the upcoming full version of the paper, we plan to extend the results using a more general set up in terms of sized digraphs. Also in Theorem 1, we imposed the condition that the set  $\mathcal{L}$  contains parts of distinct sizes. Such condition is unnecessary for unrestricted sequential structures as shown in [4]. It

might be possible to relax this condition for some classes of locally restricted sequential structures. For example, when the restriction is size-based, instead of part-based, we may allow the asymptotically free set  $\mathcal{L}$  to contain several parts of the same size. Finally it might be possible to remove the restriction that the subcomposition  $\mathbf{c}$  cannot be written in the form  $\mathbf{c} = \mathbf{xyx}$  in Theorem 2. The mapping  $\theta$  in Proposition 1 is still a bijection provided that the replacement is made at the earliest opportunity. However, the copies of  $\mathbf{c}$  in a composition may overlap here, and the maximum run length  $R(\mathbf{a})$  in a composition  $\mathbf{a} \in \mathcal{A}$  may exceed the maximum run part  $\bar{k}$  in  $\theta(\mathbf{a})$ . For example, consider the compositions  $\mathbf{c} = 121$  and  $\mathbf{a} = 12121121121$ . We note that the first two copies of  $\mathbf{c}$  overlap at the third position and  $\theta(\mathbf{a}) = 121\bar{2}$ . The maximum run part is  $\bar{2}$  here, but the maximum run length of  $\mathbf{c}$  in  $\mathbf{a}$  is 3.

## References

- [1] A.K. Agarwal,  $n$ -Colour Compositions, *Indian J. Pure Appl. Math.* **31**(11) (2000), 1421–1427.
- [2] E.A. Bender and E.R. Canfield, Locally restricted compositions II. General restrictions and infinite matrices, *Elec. J. Combin.* **16** (2009) R108, 35pp.
- [3] E.A. Bender, E.R. Canfield and Z. Gao, Locally restricted compositions IV. Nearly free large parts and gap-freeness, *Elec. J. Combin.* **19**:P14, 2012.
- [4] E.A. Bender and Z. Gao, Part Sizes of Smooth Supercritical Compositional Structures, *Combin. Prob. Comp.*, **23** (2014) 686–716.
- [5] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*, Cambridge University Press, 2009.
- [6] A. Gafni, Longest run of equal parts in a random integer composition, *Discrete Math.* **338** (2015) 236–247.
- [7] P. Grabner, A. Knopfmacher and H. Prodinger, Combinatorics of geometrically distributed random variables: Run statistics, *Theoret. Computer Sci.* **297** (2003) 261–270.
- [8] S. Heubach and T. Mansour, *Combinatorics of compositions and words*, CRC / Taylor & Francis 2009.
- [9] G. Narang and A.K. Agarwal,  $n$ -colour self-inverse compositions, *Proc. Indian Acad. Sci. Math. Sci.* **116**(3) (2006), 257–266.
- [10] G. Louchard and H. Prodinger, Probabilistic Analysis of Carlitz Compositions, *Disc. Math and Theor. Comp. Sci.* **5** (2002), 71–96.
- [11] G. Narang and A. K. Agarwal, Lattice paths and  $n$ -colour compositions, *Discrete Mathematics* **308** (2008), 1732–1740.
- [12] H.S. Wilf, The distribution of run lengths in integer compositions, *Elec. J. Combin.* **18** #2 (2011) P23, 5pp.